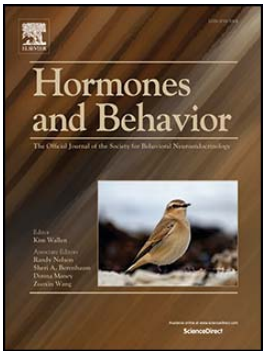


Testosterone administration does not affect men's rejections of low ultimatum game offers or aggressive mood

Carlos Cueva, R Edward Roberts, Tom J Spencer, Nisha Rani, Michelle Tempest, Philippe N Tobler, Joe Herbert, Aldo Rustichini

PII: S0018-506X(16)30072-1
DOI: doi:10.1016/j.yhbeh.2016.09.012
Reference: YHBEH 4107

To appear in: *Hormones and Behavior*

Received date: 9 February 2016
Revised date: 23 September 2016
Accepted date: 30 September 2016

Please cite this article as: Cueva, Carlos, Roberts, R Edward, Spencer, Tom J, Rani, Nisha, Tempest, Michelle, Tobler, Philippe N, Herbert, Joe, Rustichini, Aldo, Testosterone administration does not affect men's rejections of low ultimatum game offers or aggressive mood, *Hormones and Behavior* (2016), doi:10.1016/j.yhbeh.2016.09.012

Testosterone administration does not affect men's rejections of low ultimatum game offers or aggressive mood

Authors: Carlos Cueva[a*], R Edward Roberts[b], Tom J Spencer[c,d], Nisha Rani[d], Michelle Tempest[e], Philippe N Tobler[f], Joe Herbert[g], Aldo Rustichini[h,i]

Affiliations:

[a] Departamento de Fundamentos del Análisis Económico, Universidad de Alicante, Spain

[b] Division of Brain Sciences, Department of Medicine, Imperial College London, UK

[c] Department of Psychiatry, University of Cambridge, UK

[d] Cambridgeshire and Peterborough NHS Foundation Trust, Elizabeth House, Fulbourn Hospital, Cambridge, UK

[e] Royal College of Psychiatrists, UK

[f] Laboratory for Social and Neural Systems Research, Department of Economics, University of Zurich, Switzerland

[g] John van Geest Centre for Brain Repair, Department of Clinical Neurosciences, University of Cambridge, UK

[h] Department of Economics, University of Minnesota, USA

[i] Department of Physiology, Development and Neuroscience, University of Cambridge, UK

C.C. and R.E.R contributed equally to this work.

*__Contact:__ carlos.cueva@ua.es, rusti001@umn.edu.

Abstract

Correlative evidence suggests that testosterone promotes dominance and aggression. However, causal evidence is scarce and offers mixed results. To investigate this relationship, we administered testosterone for 48h to 41 healthy young adult men in a within-subjects, double-blind placebo-controlled balanced crossover design. Subjects played the role of responders in an ultimatum game, where rejecting a low offer is costly, but serves to destroy the proposer's profit. Such action can hence be interpreted as non-physical aggression in response to social provocation. In addition, subjects completed a self-assessed mood questionnaire. As expected, self-reported aggressiveness was a key predictor of ultimatum game rejections. However, while testosterone affected subjective ratings of feeling energetic and interested, our evidence strongly suggests that testosterone had no effect on ultimatum game rejections or on aggressive mood. Our findings illustrate the importance of using causal interventions to assess correlative evidence.

Keywords

Sex hormones, aggression, dominance, bargaining, neuroeconomics.

Introduction

In social groups, status hierarchies can determine access to key resources such as food, mates and territory (Saposky, 2005). Competition for status can therefore become intense and often involves aggressive challenges (Smith, 1973). Biological research points to the key role played by the hormone testosterone (T) in such interactions. For instance, a large literature indicates a positive relationship between T and aggressive or status-seeking behavior in animals (Hirschenhauser and Oliveira, 2006; Eisenegger et al, 2011). Although interspecies variation is large, animal studies have found a positive effect of exogenous T on male aggressive behavior in several species of reptiles, fish, birds and mammals (Hirschenhauser and Oliveira, 2006).

In contrast, the relationship between T and aggression or dominance in humans remains controversial despite the large body of research in this area (Archer, 1991, 2006; Mazur and Booth, 1998; Book et al, 2001; Eisenegger et al, 2011; van Honk et al, 2014). There are several important challenges to our understanding of the effects of T on dominance and aggression in humans. Firstly, there are very few studies involving direct manipulation of T in men (Kouri et al, 1995; Pope et al, 2000; O'Connor et al, 2004; Zak et al, 2009; Welling et al, 2016; Carré et al, 2016). Evidence from studies in humans is largely correlational, and, therefore, severely limited when it comes to assessing causal relationships. Secondly, the complex and multifaceted nature of human dominance and aggression makes their operationalization more difficult than in other animals (Eisenegger et al, 2011). Thirdly, due to practical and ethical difficulties in measuring actual aggressive behavior in human experiments, researchers have often relied on self-reports rather than on the measurement of actual behavior (Archer, 1991; Mazur and Booth, 1998).

In human societies, status hierarchies and dominance struggles often take place in the economic domain. In this vein, behavioral economists have become increasingly interested in the role of T in economic decision-making (Coates et al, 2010; Brañas-Garza and Rustichini, 2011). For instance, it has been suggested that T might be an important biological determinant of gender differences in negotiation, competition and risk-taking (Croson and Gneezy, 2009; Marianne, 2011). Indeed, T has been shown to be significantly correlated with financial risk-taking (Apicella et al, 2008, 2014, 2015; Sapienza et al, 2009; Stanton et al, 2011), willingness to compete (Mehta and Josephs, 2006; Carré and McCormick, 2008) and with the likelihood of pursuing a career in finance (Sapienza et al, 2009), whereas lower second-to-fourth digit ratios (2D4D) - a postulated indicator of high prenatal androgens (Manning et al, 1998) – has been found to predict long-term profits in professional traders (Coates et al, 2009). As with dominance and aggression, however, the evidence relating T to risk-taking and competition is far from conclusive. For instance, Apicella et al (2008) find a positive correlation between T and risk-taking in men, but Sapienza et al (2009) find that such correlation is only present in women, and Stanton et al (2011) observe a U-shaped relationship whereby both low and high T individuals exhibit greater risk taking. Similarly, whereas Mehta and Josephs (2006) and Carré and McCormick (2008) found that endogenously elevated T in men correlated with increased competitiveness, Apicella et al (2011) found no correlation between competitiveness and men's basal T and Mehta and Josephs (2010) found this correlation only to be present in men with low cortisol. Finally, Schipper (2015) found no relationship between men's or women's basal T and competitive bidding, a measure sensitive both to risk preferences and willingness to compete.

There are very few studies investigating the effects of exogenous manipulations of T on economic behavior, and most have focused on female, rather than male behavior

(Zethraeus et al, 2009; Bos et al, 2010; Eisenegger et al, 2010; van Honk et al, 2012; Boksem et al, 2013). Some exceptions of recent work involving T administration in healthy men include Zak et al (2009), who report suggestive evidence that it decreased generosity,[1] Wibral et al (2012) who found that it reduced lying, Carré et al (2015) who showed that it impaired socio-cognitive ability only in subjects with low 2D4D or low psychopathic traits, Cueva et al (2015) who found that it increased risk-taking in an investment task, and Carré et al (2016) who found that it led to greater aggression only in subjects scoring high in trait dominance or low in trait self-control.

Given the mixed results offered by correlational studies and the relative scarcity of experiments involving drug manipulations, it is particularly important to assess the robustness and replicability of current findings. This issue has been highlighted by recent large replication attempts in psychology and experimental economics (Open Science Collaboration, 2015; Camerer et al, 2016) as well as by recent critical reviews of oxytocin research on humans (Nave et al, 2015; Lane et al, 2016).

In this study, we use a classic experimental paradigm from behavioral economics to test the effect of T administration on aggressive bargaining behavior in men. Participants received placebo or T for 48h and, after completing a self-assessed mood questionnaire, played the role of *responders* in an ultimatum game. Subjects in this study also took part in a financial risk-taking task reported elsewhere (Cueva et al, 2015).

The ultimatum game is among the most widely used experimental paradigms on bargaining in humans (Güth et al, 1982; Güth and Kocher, 2014) and, more recently, in nonhuman primates (Jensen et al, 2007; Proctor et al, 2013). It involves two players deciding how to split a given monetary endowment. In the first stage of the game, player 1 – the *proposer* – sends an offer to player 2 – the *responder* – on how to split the endowment. In the second stage, the responder decides whether to accept or reject the offer. If the offer is accepted, each player receives the payment specified by the proposer's offer and the game ends. If the offer is rejected, both players receive *nothing* and the game ends.

The standard game theoretical prediction assuming profit maximization is that player 1 offers the smallest possible positive amount to player 2, who accepts. However, a vast body of evidence indicates that human beings seldom behave in this way when playing the ultimatum game (Güth and Kocher, 2014). Instead, players often agree on an even split of the endowment. Less generous offers are frequently rejected, indicating that many responders are willing to punish offers favoring the proposer even when it is financially costly to do so. Economists have developed various models to accommodate these findings, for instance by incorporating inequality aversion or fairness considerations into decision-makers' preferences (Fair and Schmidt, 1999; Rabin, 1993). However, ultimatum rejections have also been interpreted as a way in which individuals express anger (Xiao and Houser, 2005) and as non-physical aggression in response to social challenge (Burnham, 2007; Mehta and Beer, 2010).

A popular alternative measure in studies of aggression in humans is the Point Subtraction Aggression Paradigm (PSAP) (Pope et al, 2000; Carré and McCormick, 2008; Carré et al, 2009, 2013, 2016). In this task, subjects can sacrifice their own points in order to reduce the points of an opponent who has previously stolen points from them. The PSAP is in fact implemented as a repeated interaction against a rigged computer program. This makes its interpretation less transparent than the ultimatum game in at

[1] Their primary statistical analysis ignores the repeated-measures structure of their data and hence seriously overestimates the significance of their results.

least two ways: firstly, a repeated interaction implies that actions may be motivated by strategic considerations and not simply by an individual's aggressiveness; secondly, debriefing questionnaires reveal that participants in the PSAP often suspect that the task involves deception.

Our experiment focuses on responder behavior in the ultimatum game. According to the *challenge hypothesis*, T promotes aggression in males facing situations involving social challenge, such as competing for mates, resources or social status (Archer, 2006). Based on this view, our hypothesis is that, in men, T administration should induce more frequent rejections of low offers in the ultimatum game. We also investigated the relationship between self-reported aggressiveness and rejections of low offers, to see whether these measures were indeed related and mutually affected by T.

Methods

Subjects

All participants provided written consent to participate as approved by the Cambridge University Human Biology Research Ethics Committee. A total of 41 healthy males were recruited for the study. Three did not complete both testing sessions, resulting in a total usable sample size of 38 (mean age = 22.4 y; SD = 2.97). Participants were recruited on campus at the University of Cambridge via volunteer lists and online advertisements. To minimize the risks of possible interactions with the administration of T, a qualified clinician carried out all screening procedures, recording standard measures (blood pressure, height, and weight) and remained available throughout the experiment for medical support. Exclusion criteria included significant medical or psychiatric illness, endocrine or hormone disorders, smoking or recreational drug use.

Experimental Procedure and Drug Administration

The experiment was a within-subjects, double-blind placebo-controlled balanced crossover design. All procedures were carried out individually with each participant. Testing was divided into two blocks taking place at least one week apart. Each block involved two visits to the experimental laboratory at the University of Cambridge. Subjects were instructed not to eat or drink 30 minutes before each visit. In the first visit, all screening procedures were carried out and a baseline saliva sample was obtained. Participants were then instructed and observed applying 10g of Testogel™ (1% T gel) or a placebo of colorless hydroalcoholic gel to their shoulders and upper back, following the Testogel™ instructions. They were given a second dose of the same gel to be self-administered in 24 hours. Participants returned for testing 48 hours later and were administered a third dose of the same gel. Each participant therefore received a total of three doses of T or placebo prior to each experimental session. Additional saliva samples were collected when participants returned for the experimental session prior to the third administered dose and after participating in the behavioral tasks. In order to minimize differences in endogenous hormone levels due to diurnal variation, both experimental sessions were conducted at the same time of the day for each subject.[2] The participants did not report any side effects following the administration procedure and did not perform significantly better than chance when asked to guess in which session they received T (p = 0.618, two-sided binomial test). Behavioral testing took place after the third dose of T or placebo, between 48h and 50h after the first dose.

[2] Due to unforeseen circumstances one subject was tested in the morning one week and in the afternoon the other week.

Subjects carried out two other unrelated tasks. One of these tasks revealed a significant effect of T administration on financial risk-taking (Cueva et al, 2015).

Although the timescale of the behavioral effects of T administration has been well described in women, there is currently comparatively little data on when the behavioral effects are maximal in males (Bos et al, 2012). We therefore employed an administration procedure which would result in a sustained T increase over a 48 hour period prior to testing in order to increase the likelihood that behavioral effects would be manifested within the time window of our experimental session. Recent studies have demonstrated that a transdermal administration significantly elevates testosterone following a single dose, although the timescale for measurable changes in behavior is not well described (Eisenegger et al, 2013; Thieme et al, 2013).

Salivary T Analysis

Saliva specimens of 3ml were collected by passive drool using 12ml plastic reagent tubes (Sarstedt, UK) and immediately frozen at -80^{o}C. Samples were analyzed at the Salimetrics Centre of Excellence saliva laboratory in Cambridge (Salimetrics Europe) using a competitive immunoassay. Each assay was performed in duplicate, with inter- and intra-assay variations <6%. Of the total number of samples collected, ~7% were excluded or could not be analyzed due to either insufficient saliva volume or likely interference with the assay. In addition to this, 18% of our samples displayed T concentrations exceeding the upper concentration limit of our assay (6000pg/mL). Evidence from recent studies employing a similar topical T administration procedure have reported up to 100-fold increases in salivary T, demonstrating that this approach induces levels of circulating unbound T which can commonly exceed the range of standard assays (Schönfelder et al, 2011; Thieme et al, 2013). In these cases, we followed the approach of Schönfelder et al (2011) and set the values of these samples to 6000pg/mL. Our salivary analysis may therefore underestimate the precise magnitude of the induced increases in salivary T for some subjects.

Mood Measurement

Before participating in the ultimatum game and after receiving the third dose of T or placebo, subjects' mood was assessed using a Visual Analogue Scale (Norris, 1971). This mood scale contains a total of 16 dimensions: alert --- drowsy, calm --- excited, strong --- feeble, muzzy --- clear-headed, well-coordinated --- clumsy, lazy --- energetic, contented --- discontented, troubled --- tranquil, mentally slow --- quick-witted, tense --- relaxed, attentive --- dreamy, incompetent --- proficient, happy --- sad, aggressive --- friendly, interested --- bored, withdrawn --- outgoing.

Ultimatum Game

Proposers were recruited separately online from student lists at the University of Cambridge and did not receive T. Those who agreed to participate were asked to fill in a document and email it back before the experimental session. The document consisted of one page with simple instructions about the task and a list of possible offers from which they needed to tick one. The bottom of the page was reserved for the responders in the T administration experiment to circle their decision (accept or reject). This document was then printed out and put in an envelope. In the experimental session, participants read a similar set of instructions, this time detailing their role as responders and explaining how the offers were collected and how their responses would be communicated to the proposers (see supplementary information). The experimenter then handed an envelope with an offer, the participant circled his decision as responder

on that same sheet, put it back in the envelope and returned the envelope to the experimenter. This was repeated four times with four different offers, so each subject responded to four offers in each treatment. The four offers selected in each treatment were selected at random, but always ensuring that there was at least one 2GBP offer[3]. At the end of the experiment, one offer was randomly selected for each session and both proposer and responder were paid privately according to the particular outcome of that game.

Statistical Methods

Since our experiment employs a within-subjects design, our primary analysis is performed on paired data. The key advantage of this approach is that it allows us to factor out time-invariant individual heterogeneity. We report appropriate effect size estimations of each statistical test: Cohen's d for pairwise comparisons of continuous data, odds ratios for dichotomous data, η^2 for ANOVA, and partial η^2 and R^2 for regressions. Further details on the statistical analysis are provided in the results section.

Results

Salivary hormone analysis

As illustrated in Fig. 1a, salivary T levels were significantly elevated after 48h of T treatment compared to placebo. In order to adjust for the non-normality of the data, T levels are subsequently log-transformed for the statistical analysis. Using a repeated-measures ANOVA ($F(37, 152) = 6.80$, $p < 0.0001$, $\eta^2 = 0.95$), we found a significant time effect ($F(2,74) = 99.30$; $p < 0.0001$, $\eta^2 = 0.73$), drug effect ($F(1,36) = 186.57$; $p < 0.0001$, $\eta^2 = 0.84$) and drug-time interaction effect ($F(2,59) = 90.68$; $p < 0.0001$, $\eta^2 = 0.75$). On average, there was a 10-fold increase in salivary T levels in the treatment condition compared to placebo.[4] Post hoc t-tests show no significant time effect under placebo (0h vs 48h: $t(30) = 0.82$, $p = 0.4$; 48h vs 50h: $t(31) = 1.77$, $p = 0.08$); a significant time effect under treatment (0h vs 48h: $t(35) = 11.87$, $p < 0.0001$, Cohen's $d = 1.98$; 48h vs 50h: $t(36) = 5.43$, $p < 0.0001$, $d = 0.89$); and a significant treatment effect 48h and 50h after administration (48h: $t(33) = 11.01$, $p < 0.0001$, $d = 1.89$; 50h: $t(32) = 16.36$, $p = 0.0001$, $d = 2.86$). We also found a significant treatment effect at 0h ($t(30) = 3.70$, $p = 0.001$, $d = 0.66$), suggesting that the washout period of one week may have been insufficient to fully restore T levels back to baseline. Indeed, under placebo, subjects in the T-then-placebo condition had significantly higher T levels than subjects in the placebo-then-T condition prior to behavioral testing (0h: $t(30) = 2.47$, $p = 0.02$, $d = 0.88$; 48h: $t(33) = 3.55$, $p = 0.001$, $d = 1.20$).

In spite of the possibly insufficient washout, T levels before behavioral testing (48h) for subjects in the T-then-placebo group were still significantly higher in the T condition than in the placebo condition ($t(15) = 6.87$, $p < 0.0001$, $d = 1.73$). This difference was not significantly larger for subjects in the placebo-then-T group (difference-in-differences $t(32) = 1.18$, $p = 0.2$). In any case, we control for the possibility of a behavioral effect of insufficient washout, by employing an additional difference-in-differences analysis when estimating behavioral effects (Angrist and Pischke, 2008).

[3] Due to a temporary shortage of 2GBP offers, there was one subject who did not receive any such offers in one of the sessions.

[4] This does not imply supraphysiological T levels in blood. As shown in Thieme et al (2013) T gel administration results in very large short-run increases in salivary T but only moderate elevations of T in blood.

Ultimatum behavior

Proposers, who were recruited separately and did not receive T, were endowed with 10GBP and could offer 0, 2, 5, 8 or 10 to the responder. As expected, choices were highly clustered in either an even split or an offer to keep 8GBP and give 2GBP to the responder. Responders always accepted even splits, whereas 2GBP offers were accepted 56.7% of the time. The very few offers giving more than 5GBP to the responder were always accepted and the few offers giving nothing to the responder were always rejected. We therefore focused our analysis on responses to an offer of 2GBP.

Aggressive mood and ultimatum responses

As depicted in Fig. 1b, the self-reported mood scale administered prior to the task (Norris, 1971) indicates that feeling aggressive was a significant predictor of 2GBP ultimatum offer rejections ($t(37)= 2.43$, $p = 0.020$, fixed-effects linear regression, $F(16, 37) = 6.26$, within-$R^2 = 0.299$). The partial η^2 of aggressiveness is 0.075 and the standardized coefficient of the estimated regression is 0.20 (+1 SD in aggressiveness is associated with a 0.20 increase in the probability of rejection, 95% C.I. = [0.027, 0.37]), indicating that the correlation is fairly strong (see also Fig. 1b). This supports our hypothesis that rejections of low offers are a valid behavioral measure of an individual's inclination towards non-physical aggression.[5]

Mood and T administration

In the T condition, participants reported feeling more energetic (Wilcoxon signed-rank test, $z = 2.379$, $p = 0.017$) and interested ($z = 2.052$, $p = 0.040$), but not more aggressive ($z = -0.587$, $p = 0.6$, see Fig. 1c). This evidence is consistent with a previous placebo-controlled crossover study investigating the effects of sustained T elevations on mood, aggression and sexual behavior in young men, in which the only significant effect throughout the treatment duration was a reduction in self-reported fatigue (O'Connor et al, 2004).

T administration and ultimatum rejections

As shown in Fig. 2a, the proportion of 2GBP offers rejected was similar under both conditions: 43.4% under placebo and 43.2% under T. Of the entire group, 24 subjects did not change their responses across treatments, 6 subjects switched from accepting in the placebo treatment to rejecting in the T treatment, and 7 subjects switched from rejecting in the placebo treatment to accepting in the T treatment. An exact McNemar test, appropriate for paired binary outcomes, clearly fails to reject the null hypothesis ($\chi^2 = 0.08$, $p = 1.000$). A differences-in-differences test between the 'placebo-then-treatment' group and the 'treatment-then-placebo' group also clearly fails to reject the null hypothesis ($z = -0.161$, $p = 0.9$, Mann-Whitney U test, see Fig. 2b). Thus, we do not find any evidence that T increased the overall likelihood of rejection of low ultimatum game offers.

Although our sample size is substantially larger than most other within-subjects studies reporting behavioral effects of T administration (Bos et al, 2012), it is possible that our null findings may have been caused by insufficient power. Given that only 13 out of 38 (34%) of our subjects changed their behavior between conditions, post hoc analysis

[5] Of course, other factors such as other-regarding preferences or fairness considerations may also help to explain rejections of low ultimatum offers (Güth and Kocher, 2014).

shows that our power to detect a small, medium and large effect (odds ratios of 1.5, 3.5 and 9.0, respectively) with a one-tailed McNemar test is 0.17, 0.68 and 0.97, respectively (*G*Power*, Faul et al, 2007).

To quantify the informativeness of our null result, we estimated the 90% confidence interval of the previous exact McNemar test. The interval ranges from an odds ratio (OR) of 0.40 to 3.47. This implies that, with $p < 0.05$, we can reject an effect of T on the propensity to reject 2GBP offers of medium or larger size (OR ≥ 3.5). Alternatively, Bayes factors can be computed to estimate the odds that the null model is correct relative to alternative models that assume a positive treatment effect. Basing our prior on the positive correlations between basal T and rejection rates reported in earlier studies (Burnham, 2007; Mehta and Beer, 2010), our evidence yields Bayes factors ranging from 4.84 to 28.29, which, according to commonly used evidence categories, constitute substantial and strong evidence in favor of the null model, respectively (see supplementary information).

Beliefs and baseline T

A previous administration study reported that beliefs about having received T or placebo were significantly correlated with proposer offers in the ultimatum game (Eisenegger et al, 2010). In our data, beliefs had no significant effect on rejection rates (McNemar's $\chi^2 = 0.33$, $p = 0.774$). Controlling for beliefs, T treatment continued to have no significant effect on rejections (fixed-effects linear regression $F(2, 35) = 0.49$, within-$R^2 = 0.02$, average T effect = 0.02, 95% C.I. = (-0.15, 0.18), $t(35) = 0.2$, $p = 0.8$, see supplementary information).

Evidence of a positive correlation between baseline T levels and low offer rejections was found in two previous studies (Burnham, 2007; Mehta and Beer, 2010). We reproduced their analysis using responses in the placebo condition and found no significant difference in (log) baseline T levels between subjects who accepted 2GBP offers, and those who rejected them, although the difference was in the same direction as the previous reports ($t(30) = 0.26$, one-sided $p = 0.4$, $d = 0.10$, 95% C.I. = (-0.67, 0.86)). We repeated the analysis using placebo data from the first session only ($N = 18$) and continued to find no effect ($t(16) = 0.60$, one-sided $p = 0.3$, $d = 0.31$, 95% C.I. = (-0.77, 1.38)). Finally, Spearman's rank correlation coefficient between baseline T and rejections in this subgroup was positive but insignificant ($\rho(18) = 0.25$, $p = 0.3$). The discrepancy in our findings, however, could be due to the fact that our experiment was designed for within-subject comparisons. In particular, although each subject was tested at the same time of the day on both weeks, the times varied across subjects. Given that testosterone exhibits a marked diurnal cycle, this may have severely reduced our power to detect significant correlations between behavior and baseline T.

Discussion

Extensive evidence indicates that T is correlated with dominant behaviors aimed at asserting one's social status, often in the form of physical or non-physical aggression (Archer, 1991, 2006; Mazur and Booth, 1998; Book et al, 2001). In the context of the ultimatum game, the rejection of an unequal offer favoring the proposer has been interpreted as a form of non-physical aggression in response to social provocation (Burnham, 2007; Mehta and Beer, 2010). Thus, it is natural to formulate the hypothesis that elevated T will increase the probability of rejecting such offers.

Contrary to our hypothesis, we did not observe significant changes in responder behavior in healthy young men receiving T, compared to placebo. Controlling for beliefs about having received T or placebo did not change the interpretation of the data. Our analysis of subjects' mood prior to participation in the ultimatum game also indicates that T administration did not affect aggressiveness, which was a key predictor of ultimatum rejections. Therefore, T administration did not appear to promote mood states conducive to ultimatum rejections.

Although our study relied on a larger sample size than most previous within-subjects T administration studies (Bos et al, 2012), a post hoc power analysis shows that our study is only well-powered to detect medium to large effects. Furthermore, potential learning effects and the fact that washout duration may have been insufficient to fully restore T levels back to baseline in all participants may have also reduced our ability to detect a behavioral effect of T. On the other hand, our null result turns out to be quite informative: Bayes factors derived from our results indicate substantial support for the null hypothesis, whereas estimated confidence intervals rule out medium or larger effects of T on the probability of rejection (OR ≥ 3.5) at the 5% level.

It is unlikely that our null findings are due to a low sensitivity of the behavioral task. Crockett et al (2008) employed a similar within-subjects design and observed significant changes in ultimatum responses after acute tryptophan depletion. Furthermore, previous evidence shows that responder behavior in the ultimatum game is sensitive to subtle experimental manipulations. For example, imposing a 10 min cooling-off period before subjects can respond or allowing them to attach messages to their decisions has been shown to dramatically increase acceptance rates of low offers (Xiao and Houser, 2005; Grimm and Mengel, 2011). Even if our within-subjects design may have elicited less intra-subject variation in behavior than desirable, our evidence shows that a subject's change in his aggressive mood from one week to the next correlated with changes in his behavior in the ultimatum game. This supports the validity of our task as a sensitive behavioral measure of changes in non-physical aggression. However, none of our tests suggest an effect of T in the hypothesized direction. In addition, our experimental protocol was successful in detecting a positive effect of T on self-reported energetic mood, consistent with earlier findings (O'Connor, 2004). Similarly, the protocol induced substantial changes in financial risk-taking behavior in a separate task (Cueva et al, 2015). In particular, subjects increased their average investment in risky assets by 46% after T administration compared to placebo. Thus our experimental manipulation has been successful in modifying other forms of social and economic behavior.

There are very few studies investigating causal effects of T on bargaining behavior. A previous between-subjects study with post-menopausal women found no effect of 4 weeks of T treatment on various economic decisions including ultimatum acceptance thresholds (Zethraeus et al, 2009). Two other studies have administered T to participants in the ultimatum game with mixed evidence. The first study was a within-subjects experiment conducted with men which found that T decreased generosity (Zak et al, 2009). However, their results are only significant when considering repeated observations from each subject (N = 25) as statistically independent. Moreover, *proposer* behavior is difficult to interpret because it is likely based on the beliefs that the proposer holds regarding the responder's attitude to low offers. T could therefore induce lower offers through an increase in the tolerance to the risk of rejection. Indeed, this interpretation is supported by our recent findings regarding T and financial risk-taking (Cueva et al, 2015). Zak et al did not observe a significant effect of T on rejection thresholds. However, there are two important limitations with their evidence. Firstly, subjects stated their full strategies both as proposers and as responders at the start of

the game and before roles were assigned. Previous studies suggest that this method of elicitation (called the strategy method) and role uncertainty can have a large impact on behavior (Casari and Cason, 2009; Iriberri and Rey-Biel, 2010). Therefore, one cannot rule out that the potential effects of T might dissipate when the ultimatum game is played in this more indirect form, where subjects are invited to consider both roles in the game and specify responses to hypothetical offers. Secondly, subjects were aware of the fact that other participants were making offers after receiving T. Players' intentions and their degree of control over their actions are known to have important effects on fairness considerations and on subjects' responses to these actions (Falk et al, 2008). For this reason, folk wisdom about the effects of T on antisocial behavior (Eisenegger et al, 2010) may have easily affected responders' attitudes towards low ultimatum offers.

The second study administering T to subjects in an ultimatum game was a between-subjects experiment conducted with women which found that T increased proposer offers (Eisenegger et al, 2010). In this case, their results were only significant after controlling for beliefs about being in the placebo or T condition. They also did not observe an effect of T on responder rejections. The authors argue that an increase in fair offers and a null effect on responder behavior is consistent with the hypothesis that T promotes actions aimed at maintaining social status. The same hypothesis, however, may bring different predictions for men, since the relationship between behavior and social status is likely to be gender-specific. For instance, women cooperate more in the Prisoner's Dilemma game while being observed by their peer group whereas men cooperate less (Charness and Rustichini, 2011), suggesting important gender differences in the relationship between social status and cooperation.

A recent study implemented the PSAP in a between-subjects single T administration protocol with healthy men and found an overall insignificant effect of T on PSAP responses (Carré et al, 2016). Interestingly, a significant positive effect of T on point subtractions was found in subjects with high trait dominance or high trait self-control. Unfortunately, this study was not available until after our data collection phase so we were not able to test whether their findings would carry through in our protocol.

There are potentially other relevant interactions with T treatment that we were unable to examine. For instance, a positive association between T and dominance or competition may only be present in individuals with low basal levels of cortisol (Mehta and Josephs, 2010). Another factor that we were unable to investigate is the influence of genetic differences on an individual's sensitivity to T. For instance, the length of a polymorphic CAG repeat sequence in the androgen receptor gene is known to be inversely related to the transcriptional activity of the androgen receptor (Chamberlain et al, 1994). Finally, we must be cautious not to assume that our treatment exclusively affected T levels, since it may have also produced negative feedback effects on HPG axis activity.

The fact that we found no effect of T in this study does not rule out a T-dominance or T-aggression relationship. Numerous studies have found significant correlations between circulating T and various forms of dominant or aggressive behavior (Archer, 1991, 2006; Mazur and Booth, 1998; Book et al, 2001; Eisenegger et al, 2011; van Honk et al, 2014). One possibility is that high basal T, rather than variations in its circulating level, is a biological marker of stable individual differences in personality (Sellers et al, 2007). Another possibility is that dominance and high social status promote elevated T, rather than the other way around. Indeed, there is broad causal evidence that competition and dominance increases T but little support for the reverse effect (Archer, 2006). Finally, it is also possible that a different task involving greater social contact could elicit an effect of T on aggression or dominance. Indeed, the effects of T on behavior may be tailored for

social interaction (Bos et al, 2012), and our minimal social contact setting may have suppressed these effects. Notwithstanding these considerations, our evidence adds further support to the view that T administration does not promote aggressive, dominant or even status-seeking behavior in healthy young men.

References

Angrist, J. D., Pischke, J. S. Mostly harmless econometrics: An empiricist's companion (Princeton University Press, 2008).

Apicella, C.L., Carré, J. M., Dreber, A., 2015. Testosterone and Economic Risk Taking: A Review. Adapt. Hum. Behav. Physiol. 1–28.

Apicella, C. L., Dreber, A., Campbell, B., Gray, P. B, Hoffman, M., Little, A. C., 2008. Testosterone and financial risk preferences. Evol. Hum. Behav. 29, 384–390.

Apicella, C.L., Dreber, A., Gray, P.B., Hoffman, M., Little, A.C., Campbell, B.C., 2011. Androgens and competitiveness in men. J. Neurosci. Psychol. Econ. 4, 54–62.

Apicella, C.L., Dreber, A., Mollerstrom, J., 2014. Salivary testosterone change following monetary wins and losses predicts future financial risk-taking. Psychoneuroendocrinology 39, 58–64.

Archer, J., 1991. The influence of testosterone on human aggression. Br. J. Psychol. 82, 1–28.

Archer, J., 2006. Testosterone and human aggression: an evaluation of the challenge hypothesis. Neurosci. Biobehav. Rev. 30, 319–345.

Boksem, M. A. S., et al., 2013. Testosterone Inhibits Trust but Promotes Reciprocity. Psychol. Sci. 24, 2306–2314.

Book, A. S., Starzyk, K. B., Quinsey, V. L., 2001. The relationship between testosterone and aggression: A meta-analysis. Aggress. Violent. Behav. 6, 579–599.

Bos, P. A., Panksepp, J., Bluthé, R-M., van Honk, J., 2012. Acute effects of steroid hormones and neuropeptides on human social–emotional behavior: A review of single administration studies. Front. Neuroendocrinol. 33, 17–35.

Bos, P. A., Terburg, D., van Honk, J., 2010. Testosterone decreases trust in socially naive humans. Proc. Natl. Acad. Sci. 107, 9991–9995.

Brañas-Garza, P., Rustichini, A., 2011. Organizing effects of testosterone and economic behavior: Not just risk taking. PLoS One 6, 1–17.

Burnham, T. C., 2007. High-testosterone men reject low ultimatum game offers. Proc. Biol. Sci. 274, 2327–2330.

Camerer, C.F., Dreber, A., Forsell, E., Ho, T., Huber, J., Kirchler, M., Almenberg, J., Altmejd, A., Chan, T., Holzmeister, F., Imai, T., Isaksson, S., Nave, G., Pfeiffer, T., Razen, M., Wu, H., 2016. Evaluating replicability of laboratory experiments in economics. Science. 351, 1433–1436.

Carré, J.M., Campbell, J.A., Lozoya, E., Goetz, S.M.M., Welker, K.M., 2013. Changes in testosterone mediate the effect of winning on subsequent aggressive behaviour. Psychoneuroendocrinology 38, 2034–2041.

Carré, J.M., Geniole, S.N., Ortiz, T.L., Bird, B.M., Videto, A., Bonin, P.L., 2016. Exogenous Testosterone Rapidly Increases Aggressive Behavior in Dominant and Impulsive MenTestosterone modulates aggression in dominant/impulsive men. Biol. Psychiatry 1–8.

Carré, J.M., McCormick, C.M., 2008. Aggressive behavior and change in salivary testosterone concentrations predict willingness to engage in a competitive task. Hormones and Behavior 54, 403–409.

Carré, J.M., Ortiz, T.L., Labine, B., Moreau, B.J.P., Viding, E., Neumann, C.S., Goldfarb, B., 2015. Digit ratio (2D:4D) and psychopathic traits moderate the effect of exogenous testosterone on socio-cognitive processes in men. Psychoneuroendocrinology 62, 319–326.

Carré, J. M., Putnam, S. K., McCormick, C. M., 2009. Testosterone responses to competition predict future aggressive behaviour at a cost to reward in men. Psychoneuroendocrinology 34, 561–570.

Casari, M., Cason, T. N., 2009. The strategy method lowers measured trustworthy behavior. Econ. Lett. 103, 157–159.

Chamberlain, N. L., Driver, E. D., Miesfeld, R. L., 1994. The length and location of CAG trinucleotide repeats in the androgen receptor N-terminal domain affect transactivation function. Nucleic Acids Res. 22, 3181–3186.

Charness, G., Rustichini, A., 2011. Gender differences in cooperation with group membership. Games Econ. Behav. 72, 77–85.

Coates, J. M., Gurnell, M., Rustichini, A., 2009. Second-to-fourth digit ratio predicts success among high-frequency financial traders. Proc. Natl. Acad. Sci. 106, 623–628.

Coates, J. M., Gurnell, M., Sarnyai, Z., 2010. From molecule to market: steroid hormones and financial risk-taking. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 365, 331–343.

Crockett, M. J., Clark, L., Tabibnia, G., Lieberman, M.D., Robbins, T.W., 2008. Serotonin modulates behavioral reactions to unfairness. Science 320, 1739.

Croson, R., Gneezy, U., 2009. Gender Differences in Preferences. J. Econ. Lit. 47, 448–474.

Cueva, C. et al., 2015. Cortisol and testosterone increase financial risk taking and may destabilize markets. Sci. Rep. 5.

Eisenegger, C., von Eckardstein, A., Fehr, E. and von Eckardstein, S., 2013. Pharmacokinetics of testosterone and estradiol gel preparations in healthy young men. Psychoneuroendocrinology 38, 171–8.

Eisenegger, C., Haushofer, J., Fehr, E., 2011. The role of testosterone in social interaction. Trends Cogn. Sci. 15, 263–271.

Eisenegger, C., Naef, M., Snozzi, R., Heinrichs, M., Fehr, E., 2010. Prejudice and truth about the effect of testosterone on human bargaining behaviour. Nature 463, 356–359.

Falk, A., Fehr, E., Fischbacher, U., 2008. Testing theories of fairness—Intentions matter. Games Econ. Behav. 62, 287-303.

Faul, F., Erdfelder, E., Lang, A.-G., Buchner, A., 2007. G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behavior Research Methods 39, 175-191.

Grimm, V., Mengel, F., 2011. Let me sleep on it: Delay reduces rejection rates in ultimatum games. Econ. Lett. 111, 113–115.

Güth, W., Kocher, M. G., 2014. More than thirty years of ultimatum bargaining experiments: Motives, variations, and a survey of the recent literature. J. Econ. Behav. Organ. 108, 396–409.

Güth, W., Schmittberger, R., Schwarze, B., 1982. An experimental analysis of ultimatum bargaining. J. Econ. Behav. Organ. 3, 367-388.

Hirschenhauser, K., Oliveira, R. F., 2006. Social modulation of androgens in male vertebrates: Meta-analyses of the challenge hypothesis. Anim. Behav. 71, 265–277.

Iriberri, N., Rey-Biel, P., 2010. The role of role uncertainty in modified dictator games. Exp. Econ. 14, 160–180.

Jensen, K., Call, J., Tomasello, M., 2007. Chimpanzees are rational maximizers in an ultimatum game. Science 318, 107–109.

Kouri, E.M. et al., 1995. Increased aggressive responding in male volunteers following the administration of gradually increasing doses of testosterone cypionate. Drug and alcohol dependence, 40, 73–9.

Lane, A., Luminet, O., Nave, G., Mikolajczak, M., 2016. Is there a publication bias in behavioral intranasal oxytocin research on humans? Opening the file drawer of one lab. J. Neuroendocrinol. 28.

Manning, J. T., Scutt, D., Wilson, J., Lewis-Jones, D. I., 1998. The ratio of 2nd to 4th digit length: a predictor of sperm numbers and concentrations of testosterone, luteinizing hormone and oestrogen. Hum. Reprod. 13, 3000–3004.

Marianne, B., 2011. New perspectives on gender. Hand. Labor Econ. 4, 1543–1590.

Mazur, A., Booth, A., 1998. Testosterone and dominance in men. Behav. Brain Sci. 21, 353–63.

Mehta, P. H., Beer, J., 2010. Neural mechanisms of the testosterone-aggression relation: the role of orbitofrontal cortex. J. Cogn. Neurosci. 22, 2357–2368.

Mehta, P.H., Josephs, R.A., 2006. Testosterone change after losing predicts the decision to compete again. Hormones and Behavior, 50, 684–692.

Mehta, P.H., Josephs, R. A, 2010. Testosterone and cortisol jointly regulate dominance: evidence for a dual-hormone hypothesis. Horm. Behav. 58, 898–906.

Nave, G., Camerer, C., McCullough, M., 2015. Does oxytocin increase trust in humans? A critical review of research. Perspect. Psychol. Sci. 10, 772–789.

Norris ,H., 1971. The action of sedatives on brainstem oculomotor systems in man. Neuropharmacology 10, 181–191.

O'Connor, D. B., Archer, J., Wu, F. C. W., 2004. Effects of Testosterone on Mood, Aggression, and Sexual Behavior in Young Men: A Double-Blind, Placebo-Controlled, Cross-Over Study. J. Clin. Endocrinol. Metab. 89, 2837–2845.

Open Science Collaboration, 2015. Estimating the reproducibility of psychological science. Science 349, aac4716.

Pope, H. G., Kouri, E. M., Hudson, J. I., 2000. Effects of supraphysiologic doses of testosterone on mood and aggression in normal men: a randomized controlled trial. Arch. Gen. Psychiatry 57, 133–140.

Proctor, D., Williamson, R. A., de Waal, F. B., Brosnan, S. F., 2007. Chimpanzees play the ultimatum game. Proc. Natl. Acad. Sci. 110, 2070–2075.

Sapienza, P., Zingales, L., Maestripieri, D., 2009. Gender differences in financial risk aversion and career choices are affected by testosterone. Proc. Natl. Acad. Sci. 106, 15268–15273.

Sapolsky, R. M., 2005. The Influence of Social Hierarchy on Primate Health. Science 308, 648–652.

Schipper, B.C., 2015. Sex Hormones and Competitive Bidding. Manage. Sci. 61, 249–266.

Schönfelder, M., Hofmann, H., Anielski, P., Thieme, D., Oberhoffer, R., Michna, H., 2011. Gene expression profiling in human whole blood samples after controlled testosterone application and exercise. Drug Test. Anal. 3, 652–660.

Sellers, J. G., Mehl, M. R., Josephs, R. A., 2007. Hormones and personality: Testosterone as a marker of individual differences. J. Res. Pers. 41, 126–138.

Smith, J. M., Price, G. R., 1973. The Logic of Animal Conflict. Nature, 246, 15.

Stanton, S. J., Mullette-Gillman, O. A., McLaurin, R. E., Kuhn, C. M., LaBar, K. S., Platt, M. L., Huettel, S. A., 2011. Low- and high-testosterone individuals exhibit decreased aversion to economic risk. Psychol. Sci. 22, 447–53.

Thieme, D., Rautenberg, C., Grosse, J., Schoenfelder, M., 2013. Significant increase of salivary testosterone levels after single therapeutic transdermal administration of testosterone: suitability as a potential screening parameter in doping control. Drug Test. Anal. 5, 819–825.

van Honk, J., Bos, P. A., Terburg, D., 2014. Testosterone and Dominance in Humans: Behavioral and Brain Mechanisms. In New Frontiers in Social Neuroscience (pp. 201-214). Springer International Publishing.

van Honk, J., Montoya, E. R., Bos, P. A., van Vugt, M., Terburg, D., 2012. New evidence on testosterone and cooperation. Nature 485, E4–5.

Welling, L. M., Moreau, B. P., Bird, B. M., Hansen, S., Carré, J. M., 2016. Exogenous testosterone increases men's perceptions of their own physical dominance. Psychoneuroendocrinology 64, 136–142.

Wibral, M., Dohmen, T., Klingmu, D., Weber, B., Falk, A., 2012. Testosterone Administration Reduces Lying in Men ¨. PLoS One 7, e46774.

Xiao, E., Houser, D., 2005. Emotion expression in human punishment behavior. Proc. Natl. Acad. Sci. 102, 7398–7401.

Zak, P. J. et al., 2009. Testosterone administration decreases generosity in the ultimatum game. PLoS One 4, e8330.

Zethraeus, N., et al., 2009. A randomized trial of the effect of estrogen and testosterone on economic behavior. Proc. Natl. Acad. Sci. 106, 6535–6538.

Acknowledgements
We thank Paul Fletcher, Emilio Fernandez-Egea, Ashley Grossman, Helen Gossage and Sheila Skidmore for their help during the project. This work was supported by the Economic and Social Research Council (grant ES/G005230/1). CC was supported by the Spanish Ministerio de Economía y Competitividad (ECO2012-34928). PNT was supported by the Swiss National Science Foundation (PP00P1_128574 PP00P1_150739 and CRSII3_141965). RER was supported by the UK Medical Research Council (Mr/J004685/1).

Author contributions
C.C., R.E.R., A.R., J.H. and P.N.T., designed research; C.C. and R.E.R., carried out research, T.S., N.R., M.T. and J.H. assisted in research, C.C., R.E.R., A.R., J.H. and P.N.T. analyzed data and wrote the paper.

Competing interests
We have no competing financial interests

Figure 1. (A) Mean salivary testosterone concentrations (± 1 SE) in the placebo and testosterone treatments. Subjects received 3 daily consecutive doses of testosterone or placebo each week. Saliva was assayed immediately before the first and last dose (0h and 48h). Behavioral testing took place at 49h. **(B)** Aggressive mood and ultimatum responses to unfavorable offers. Subjects who were more likely to reject 2GBP offers in week 2 than in week 1 were also feeling more aggressive in week 2 than in week 1, according to their responses to the visual analogue mood scale (t = 2.43, p = 0.020, see table S2). **(C)** Testosterone treatment and aggressive mood. Subjects did not report feeling more aggressive in the testosterone condition than in the placebo condition (z = -0.587, p = 0.6, see table S3).

Figure 2. (A) Testosterone treatment and proportion of unfavorable offers accepted. Subjects were not significantly more likely to accept 2GBP offers in the placebo condition than in the testosterone condition (McNemar's, χ^2 = 0.08, exact significance p = 1.000). **(B)** Treatment order had no effect on the proportion of unfavorable offers accepted in each week (z = -0.161, p = 0.9).

Supplementary information
Supplementary information accompanies this paper at http://www... (available upon publication)

Figure 1

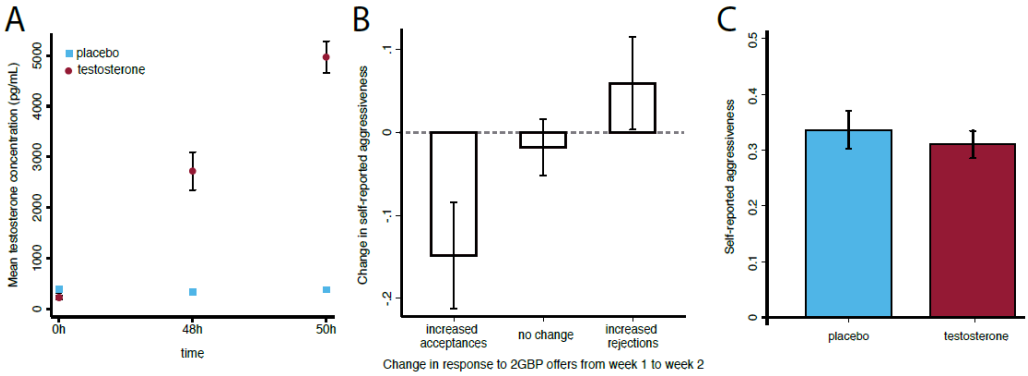
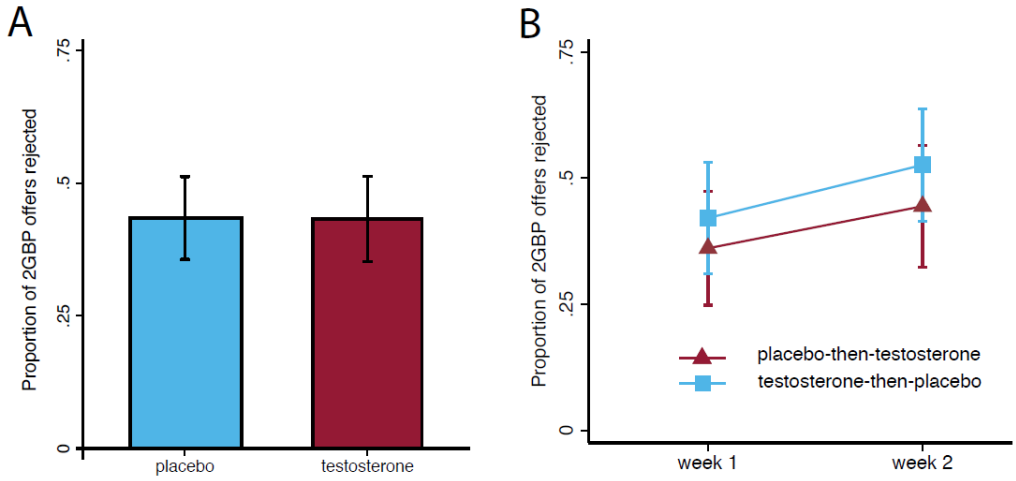


Figure 2



Highlights

- 41 healthy young men receive T or placebo in a within-subjects double-blind design
- Subjects respond to unfavorable offers in an incentivized ultimatum game
- Other measures include self-reported mood and baseline T
- Aggressive mood predicts ultimatum game rejections
- T increases energetic mood but not aggressiveness or ultimatum game rejections